

#5
PCT09

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:08

Input Set : N:\Crif3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

ENTERED

PS

```

1 <110> APPLICANT: Zavada, Jan
2   Pastorekova, Silvia
3   Pastorek, Jaromir
4 <120> TITLE OF INVENTION: MN Gene and Protein
5 <130> FILE REFERENCE: D-0021.5 PCT
6 <140> CURRENT APPLICATION NUMBER: US/09/807,949A
7 <141> CURRENT FILING DATE: 2001-08-09
8 <150> PRIOR APPLICATION NUMBER: 09/177,776
9 <151> PRIOR FILING DATE: 1998-10-23
10 <150> PRIOR APPLICATION NUMBER: 09/178,115
11 <151> PRIOR FILING DATE: 1998-10-23
12 <160> NUMBER OF SEQ ID NOS: 143
13 <170> SOFTWARE: PatentIn Ver. 2.1
15 <210> SEQ ID NO: 1
16 <211> LENGTH: 1522
17 <212> TYPE: DNA
18 <213> ORGANISM: HUMAN
19 <220> FEATURE:
20 <221> NAME/KEY: CDS
21 <222> LOCATION: (13)..(1389)
22 <221> NAME/KEY: mat_peptide
23 <222> LOCATION: (124)..(1389)
24 <400> SEQUENCE: 1
25   acagtcagcc gc atg gct ccc ctg tgc ccc agc ccc tgg ctc cct ctg ttg 51
26               Met Ala Pro Leu Cys Pro Ser Pro Trp Leu Pro Leu Leu
27               -35               -30               -25
28   atc ccg gcc cct gct cca ggc ctc act gtg caa ctg ctg ctg tca ctg 99
29   Ile Pro Ala Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu
30               -20               -15               -10
31   ctg ctt ctg atg cct gtc cat ccc cag agg ttg ccc cgg atg cag gag 147
32   Leu Leu Leu Met Pro Val His Pro Gln Arg Leu Pro Arg Met Gln Glu
33               -5               -1   1               5
34   gat tcc ccc ttg gga gga ggc tct tct ggg gaa gat gac cca ctg ggc 195
35   Asp Ser Pro Leu Gly Gly Gly Ser Ser Gly Glu Asp Asp Pro Leu Gly
36       10               15               20
37   gag gag gat ctg ccc agt gaa gag gat tca ccc aga gag gag gat cca 243
38   Glu Glu Asp Leu Pro Ser Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro
39       25               30               35               40
40   ccc gga gag gag gat cta cct gga gag gag gat cta cct gga gag gag 291
41   Pro Gly Glu Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro Gly Glu Glu
42               45               50               55
43   gat cta cct gaa gtt aag cct aaa tca gaa gaa gag ggc tcc ctg aag 339
44   Asp Leu Pro Glu Val Lys Pro Lys Ser Glu Glu Glu Gly Ser Leu Lys
45               60               65               70
46   tta gag gat cta cct act gtt gag gct cct gga gat cct caa gaa ccc 387
47   Leu Glu Asp Leu Pro Thr Val Glu Ala Pro Gly Asp Pro Gln Glu Pro
48       75               80               85

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:08

Input Set : N:\Crif3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

49	cag aat aat gcc cac agg gac aaa gaa ggg gat gac cag agt cat tgg	435
50	Gln Asn Asn Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp	
51	90 95 100	
52	cgc tat gga ggc gac ccg ccc tgg ccc cgg gtg tcc cca gcc tgc gcg	483
53	Arg Tyr Gly Gly Asp Pro Pro Trp Pro Arg Val Ser Pro Ala Cys Ala	
54	105 110 115 120	
55	ggc cgc ttc cag tcc ccg gtg gat atc cgc ccc cag ctc gcc gcc ttc	531
56	Gly Arg Phe Gln Ser Pro Val Asp Ile Arg Pro Gln Leu Ala Ala Phe	
57	125 130 135	
58	tgc ccg gcc ctg cgc ccc ctg gaa ctc ctg ggc ttc cag ctc ccg ccg	579
59	Cys Pro Ala Leu Arg Pro Leu Glu Leu Gly Phe Gln Leu Pro Pro	
60	140 145 150	
61	ctc cca gaa ctg cgc ctg cgc aac aat ggc cac agt gtg caa ctg acc	627
62	Leu Pro Glu Leu Arg Leu Arg Asn Asn Gly His Ser Val Gln Leu Thr	
63	155 160 165	
64	ctg cct cct ggg cta gag atg gct ctg ggt ccc ggg cgg gag tac cgg	675
65	Leu Pro Pro Gly Leu Glu Met Ala Leu Gly Pro Gly Arg Glu Tyr Arg	
66	170 175 180	
67	gct ctg cag ctg cat ctg cac tgg ggg gct gca ggt cgt ccg ggc tcg	723
68	Ala Leu Gln Leu His Leu His Trp Gly Ala Ala Gly Arg Pro Gly Ser	
69	185 190 195 200	
70	gag cac act gtg gaa ggc cac cgt ttc cct gcc gag atc cac gtg gtt	771
71	Glu His Thr Val Glu Gly His Arg Phe Pro Ala Glu Ile His Val Val	
72	205 210 215	
73	cac ctc agc acc gcc ttt gcc aga gtt gac gag gcc ttg ggg cgc ccg	819
74	His Leu Ser Thr Ala Phe Ala Arg Val Asp Glu Ala Leu Gly Arg Pro	
75	220 225 230	
76	gga ggc ctg gcc gtg ttg gcc gcc ttt ctg gag gag ggc ccg gaa gaa	867
77	Gly Gly Leu Ala Val Leu Ala Ala Phe Leu Glu Glu Gly Pro Glu Glu	
78	235 240 245	
79	aac agt gcc tat gag cag ttg ctg tct cgc ttg gaa gaa atc gct gag	915
80	Asn Ser Ala Tyr Glu Gln Leu Leu Ser Arg Leu Glu Glu Ile Ala Glu	
81	250 255 260	
82	gaa ggc tca gag act cag gtc cca gga ctg gac ata tct gca ctc ctg	963
83	Glu Gly Ser Glu Thr Gln Val Pro Gly Leu Asp Ile Ser Ala Leu Leu	
84	265 270 275 280	
85	ccc tct gac ttc agc cgc tac ttc caa tat gag ggg tct ctg act aca	1011
86	Pro Ser Asp Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr Thr	
87	285 290 295	
88	ccg ccc tgt gcc cag ggt gtc atc tgg act gtg ttt aac cag aca gtg	1059
89	Pro Pro Cys Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr Val	
90	300 305 310	
91	atg ctg agt gct aag cag ctc cac acc ctc tct gac acc ctg tgg gga	1107
92	Met Leu Ser Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp Gly	
93	315 320 325	
94	cct ggt gac tct cgg cta cag ctg aac ttc cga gcg acg cag cct ttg	1155
95	Pro Gly Asp Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro Leu	
96	330 335 340	
97	aat ggg cga gtg att gag gcc tcc ttc cct gct gga gtg gac agc agt	1203

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:08

Input Set : N:\Crf3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

```

98      Asn Gly Arg Val Ile Glu Ala Ser Phe Pro Ala Gly Val Asp Ser Ser
99      345              350              355              360
100     cct cgg gct gct gag cca gtc cag ctg aat tcc tgc ctg gct gct ggt 1251
101     Pro Arg Ala Ala Glu Pro Val Gln Leu Asn Ser Cys Leu Ala Ala Gly
102              365              370              375
103     gac atc cta gcc ctg gtt ttt ggc ctc ctt ttt gct gtc acc agc gtc 1299
104     Asp Ile Leu Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val
105              380              385              390
106     gcg ttc ctt gtg cag atg aga agg cag cac aga agg gga acc aaa ggg 1347
107     Ala Phe Leu Val Gln Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly
108              395              400              405
109     ggt gtg agc tac cgc cca gca gag gta gcc gag act gga gcc 1389
110     Gly Val Ser Tyr Arg Pro Ala Glu Val Ala Glu Thr Gly Ala
111     410              415              420
112     tagaggctgg atcttgagaga atgtgagaag ccagccagag gcactctgagg gggagccggt 1449
113     aactgtcctg tctgtctcat tatgccactt ccttttaact gccaaagaaat tttttaaaat 1509
114     aaatatattat aat 1522
116 <210> SEQ ID NO: 2
117 <211> LENGTH: 459
118 <212> TYPE: PRT
119 <213> ORGANISM: HUMAN
120 <400> SEQUENCE: 2
121     Met Ala Pro Leu Cys Pro Ser Pro Trp Leu Pro Leu Leu Ile Pro Ala
122              -35              -30              -25
123     Pro Ala Pro Gly Leu Thr Val Gln Leu Leu Leu Ser Leu Leu Leu Leu
124              -20              -15              -10
125     Met Pro Val His Pro Gln Arg Leu Pro Arg Met Gln Glu Asp Ser Pro
126     -5              -1 1              5              10
127     Leu Gly Gly Gly Ser Ser Gly Glu Asp Asp Pro Leu Gly Glu Glu Asp
128              15              20              25
129     Leu Pro Ser Glu Glu Asp Ser Pro Arg Glu Glu Asp Pro Pro Gly Glu
130              30              35              40
131     Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro Gly Glu Glu Asp Leu Pro
132              45              50              55
133     Glu Val Lys Pro Lys Ser Glu Glu Glu Gly Ser Leu Lys Leu Glu Asp
134              60              65              70              75
135     Leu Pro Thr Val Glu Ala Pro Gly Asp Pro Gln Glu Pro Gln Asn Asn
136              80              85              90
137     Ala His Arg Asp Lys Glu Gly Asp Asp Gln Ser His Trp Arg Tyr Gly
138              95              100              105
139     Gly Asp Pro Pro Trp Pro Arg Val Ser Pro Ala Cys Ala Gly Arg Phe
140              110              115              120
141     Gln Ser Pro Val Asp Ile Arg Pro Gln Leu Ala Ala Phe Cys Pro Ala
142              125              130              135
143     Leu Arg Pro Leu Glu Leu Leu Gly Phe Gln Leu Pro Pro Leu Pro Glu
144              140              145              150              155
145     Leu Arg Leu Arg Asn Asn Gly His Ser Val Gln Leu Thr Leu Pro Pro
146              160              165              170
147     Gly Leu Glu Met Ala Leu Gly Pro Gly Arg Glu Tyr Arg Ala Leu Gln

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:08

Input Set : N:\Crf3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

```

148          175          180          185
149  Leu His Leu His Trp Gly Ala Ala Gly Arg Pro Gly Ser Glu His Thr
150          190          195          200
151  Val Glu Gly His Arg Phe Pro Ala Glu Ile His Val Val His Leu Ser
152          205          210          215
153  Thr Ala Phe Ala Arg Val Asp Glu Ala Leu Gly Arg Pro Gly Gly Leu
154  220          225          230          235
155  Ala Val Leu Ala Ala Phe Leu Glu Glu Gly Pro Glu Glu Asn Ser Ala
156          240          245          250
157  Tyr Glu Gln Leu Leu Ser Arg Leu Glu Glu Ile Ala Glu Glu Gly Ser
158          255          260          265
159  Glu Thr Gln Val Pro Gly Leu Asp Ile Ser Ala Leu Leu Pro Ser Asp
160          270          275          280
161  Phe Ser Arg Tyr Phe Gln Tyr Glu Gly Ser Leu Thr Thr Pro Pro Cys
162          285          290          295
163  Ala Gln Gly Val Ile Trp Thr Val Phe Asn Gln Thr Val Met Leu Ser
164  300          305          310          315
165  Ala Lys Gln Leu His Thr Leu Ser Asp Thr Leu Trp Gly Pro Gly Asp
166          320          325          330
167  Ser Arg Leu Gln Leu Asn Phe Arg Ala Thr Gln Pro Leu Asn Gly Arg
168          335          340          345
169  Val Ile Glu Ala Ser Phe Pro Ala Gly Val Asp Ser Ser Pro Arg Ala
170          350          355          360
171  Ala Glu Pro Val Gln Leu Asn Ser Cys Leu Ala Ala Gly Asp Ile Leu
172          365          370          375
173  Ala Leu Val Phe Gly Leu Leu Phe Ala Val Thr Ser Val Ala Phe Leu
174  380          385          390          395
175  Val Gln Met Arg Arg Gln His Arg Arg Gly Thr Lys Gly Gly Val Ser
176          400          405          410
177  Tyr Arg Pro Ala Glu Val Ala Glu Thr Gly Ala
178          415          420
180 <210> SEQ ID NO: 3
181 <211> LENGTH: 29
182 <212> TYPE: DNA
183 <213> ORGANISM: HUMAN
184 <400> SEQUENCE: 3
185  cgcccagtggtgtcatcttcc ccagaagag 29
187 <210> SEQ ID NO: 4
188 <211> LENGTH: 19
189 <212> TYPE: DNA
190 <213> ORGANISM: HUMAN
191 <400> SEQUENCE: 4
192  ggaatccctcc tgcattccgg 19
194 <210> SEQ ID NO: 5
195 <211> LENGTH: 10898
196 <212> TYPE: DNA
197 <213> ORGANISM: HUMAN
198 <220> FEATURE:
199 <221> NAME/KEY: gene

```

RAW SEQUENCE LISTING

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:08

Input Set : N:\Cr3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

200 <222> LOCATION: (1)..(10898)

201 <223> OTHER INFORMATION: full-length MN genomic sequence

W--> 202 <221> NAME/KEY: unsure of base at position 1974

203 <222> LOCATION: (1974)

204 <223> OTHER INFORMATION: unsure of base at position 1974, which is in the 5' region flanking the transcription initiation site (3507) as determined by RNase protection assay.

205

206 <400> SEQUENCE: 5

207	ggatcctgtt	gactcgtgac	cttaccceca	accctgtgct	ctctgaaaca	tgagctgtgt	60
208	ccactcaggg	ttaaattgga	taagggcggt	gcaagatgtg	ctttgttaaa	cagatgcttg	120
209	aaggcagcat	gctcgtaaag	agtcatacacc	aatccctaata	ctcaagtaat	cagggacaca	180
210	aacactgcgg	aaggccgcag	ggtcctctgc	ctaggaaaac	cagagacctt	tggtcacttg	240
211	tttatctgac	cttccctcca	ctattgtcca	tgaccctgcc	aaatccccct	ctgtgagaaa	300
212	cacccaagaa	ttatcaataa	aaaaataaat	ttaaaaaaaa	aatacaaaaa	aaaaaaaaaa	360
213	aaaaaaaaaa	gacttacgaa	tagttattga	taaatgaata	gctattggta	aagccaagta	420
214	aatgatcata	ttcaaaacca	gacggccatc	atcacagctc	aagtctacct	gatttgatct	480
215	ctttatcatt	gtcattcttt	ggattcacta	gattagtcac	catcctcaaa	attctcccc	540
216	aagttctaata	tacgttccaa	acatttaggg	gttacatgaa	gcttgaacct	actaccttct	600
217	ttgcttttga	gccatgagtt	gtaggaatga	tgagtttaca	ccttacatgc	tggggattaa	660
218	tttaaaacttt	acctctaagt	cagttgggta	gcctttggct	tatttttgta	gctaattttg	720
219	tagttaatgg	atgcaactgt	aatcttgcta	tgatagtttt	cctccacact	ttgccactag	780
220	gggtaggtag	gtactcagtt	ttcagtaatt	gcttacctaa	gaccctaagc	cctattttctc	840
221	ttgtactggc	ctttatctgt	aatatgggca	tatttaatac	aatataattt	ttggagtttt	900
222	tttgtttgtt	tggtttgttg	tttttttgag	acggagtctt	gcactctgtc	tgcccaggct	960
223	ggagtagcag	tggtgccatc	tcggtcact	gcaagctcca	cctcccaggt	tcacgccatt	1020
224	ttcctgcctc	agcctcccga	gtagctggga	ctacaggcgc	ccgccaccat	gcccggctaa	1080
225	ttttttgtat	ttttggtaga	gacgggggtt	caccgtgtta	gccagaatgg	tctcgatctc	1140
226	ctgacttcgt	gatccacccg	cctcggcctc	ccaaagtctt	gggattacag	gtgtgagcca	1200
227	ccgcacctgg	ccaatttttt	gagtccttta	aagtaaaaaa	atgtcttgta	agctggtaac	1260
228	tatggtacat	ttccttttat	taatgtggty	ctgacggtca	tataggttct	tttgagtttg	1320
229	gcactgcat	gctacttttt	gcagtccttt	cattacattt	ttctctcttc	atttgaagag	1380
230	catgttatat	cttttagctt	caottggctt	aaaagggtct	ctcattagcc	taacacagtg	1440
231	tcattgttgg	taccacttgg	atcataagtg	gaaaaacagt	caagaaattg	cacagtaata	1500
232	cttgtttgta	agagggatga	ttcaggtgaa	tctgacacta	agaaactccc	ctacctgagg	1560
233	tctgagattc	ctctgacatt	gctgtatata	ggcttttcct	ttgacagcct	gtgactgcgg	1620
234	actatttttc	ttaagcaaga	tatgctaaag	ttttgtgagc	ctttttccag	agagaggtct	1680
235	catacttgca	tcaagtgaga	acatataatg	tctgcatgtt	tccatatttc	aggaatgttt	1740
236	gcttgtgttt	tatgctttta	tatagacagg	gaaacttggt	cctcagtgac	ccaaaagagg	1800
237	tggggaattgt	tattggatat	catcattggc	ccacgctttc	tgaccttgga	aacaattaaag	1860
238	gggtcataat	ctcaattctg	tcagaattgg	tacaagaaat	agctgctatg	tttcttgaca	1920
W--> 239	ttccacttgg	taggaaataa	gaatgtgaaa	ctcttcagtt	ggtgtgtgtc	cctngttttt	1980
240	ttgcaatttc	cttcttactg	tgtaaaaaaa	aagtatgatc	ttgctctgag	aggtgaggca	2040
241	ttcttaatac	tgatctttta	agatcaataa	tataatcctt	tcaaggatta	tgtctttatt	2100
242	ataataaaga	taatttgtct	ttaacagaat	caataatata	atcccttaaa	ggattatata	2160
243	tttgctgggc	gcagtggctc	acacctgtaa	tcccagcact	ttgggtggcc	aaggtggaag	2220
244	gatcaaattt	gcctacttct	atattatctt	ctaaagcaga	attcatctct	cttccctcaa	2280
245	tatgatgata	ttgacagggt	ttgcctcacc	tcactagatt	gtgagctcct	gctcagggca	2340
246	ggtagcgttt	tttgtttttg	ttttgtttt	tcttttttga	gacagggtct	tgctctgtca	2400
247	ccagggccag	agtgcattgg	tacagtctca	gctcactgca	gcctcaaccg	cctcggctca	2460
248	aaccatcatc	ccatttcagc	ctcctgagta	gctgggacta	caggcacatg	ccattacacc	2520

Use of n and/or Xaa has been detected in the Sequence Listing.
Review the Sequence Listing to insure a corresponding explanation is presented in the <220> to <223> fields of each sequence using n or Xaa.

VERIFICATION SUMMARY

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:09

Input Set : N:\Crf3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

L:202 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:5
L:239 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:424 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:9
L:569 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:25
L:574 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25
L:586 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:26
L:591 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26
L:1119 M:259 W: Allowed number of lines exceeded, <223> Other Information:
L:1120 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:58
L:1126 M:259 W: Allowed number of lines exceeded, <223> Other Information:
L:1160 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:58
L:1469 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:90
L:1475 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:90
L:1516 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:90
L:1723 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:110
L:1740 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:110

STATISTICS SUMMARY

DATE: 10/17/2001

PATENT APPLICATION: US/09/807,949A

TIME: 12:15:09

Input Set : N:\Crf3\10162001\I807949.raw

Output Set: N:\CRF3\10172001\I807949A.raw

Application Serial Number: US/09/807,949A

Alpha or Numeric: Numeric

Application Class:

Application File Date: 08-09-2001

Art Unit: PCT09

Software Application: PatentIn

Total Number of Sequences: 143

Total Nucleotides: 33606

Total Amino Acids: 1828

Number of Errors: 0

Number of Warnings: 16

Number of Corrections: 0

MESSAGE SUMMARY

257 W: 8 (Feature value mis-spelled or invalid)

259 W: 2 (Allowed number of lines exceeded)

341 W: 6 ((46) "n" or "Xaa" used)